

Autour du nom propre conceptuel

Introduction

Cette partie est entièrement consacrée à décrire en détail tous les concepts clés de la modélisation des noms propres et les relations qu’entretiennent ces différents concepts entre eux. Ces concepts nous ont finalement permis de proposer une modélisation sous la forme d’un graphe orienté structuré sur plusieurs niveaux, pouvant parfois être très complexe en fonction de la langue que l’on souhaite intégrer à notre dictionnaire électronique.

C’est à partir des différents travaux réalisés en TAL et en linguistique sur les phénomènes dérivationnels et morphologiques qui se rapportent aux noms propres dans plusieurs langues, telles que le français, l’anglais, l’allemand et particulièrement le serbe, et aux relations que les noms propres entretiennent entre eux, que nous avons pu définir ces concepts. Le serbe constitue sans doute une des langues européennes les plus riches au niveau dérivationnel et au niveau morphologique que nous ayons pu étudier au cours de nos travaux.

Nous avons défini deux principaux concepts : le nom propre conceptuel et le prolexème. Autour de ces deux concepts, nous avons défini d’autres concepts, des relations qui ne dépendent pas des langues et des relations qui dépendent de la langue.

Nous avons ensuite regroupé nos différents concepts et relations du domaine des noms propres sous la forme d’une arborescence qui peut se décomposer en quatre niveaux distincts :

- le niveau méta-conceptuel : la typologie et l’existence, que nous présenterons dans le prochain chapitre.
- le niveau conceptuel : le nom propre conceptuel et les relations qui ne dépendent pas de la langue.
- le niveau linguistique : le prolexème, les alias, les dérivés et les relations qui dépendent de la langue.
- le niveau des instances : l’ensemble des formes fléchies d’une langue.

Dans les débuts de nos travaux [Tran et al., 2004], nous avons créé notre premier modèle des noms propres en nous basant uniquement sur les langues française, anglaise et allemande.

Nous nous sommes vite rendu compte de la limite du modèle que nous proposons en travaillant avec Duško Vitas et Cvetana Krstev, dans le cadre du projet Égide Pavle-Savic, sur l’intégration du serbe dans notre dictionnaire électronique de noms propres. Ce modèle, basé essentiellement sur ces trois langues, était bien trop simple et ne prenait pas en compte les mécanismes très compliqués de dérivation et de morphologie produits par la langue serbe. Un voyage à Belgrade, à la fois inoubliable et extrêmement enrichissant, nous a permis de remettre en cause notre modèle et de le perfectionner en rajoutant de nouveaux concepts

[Krstev et al., 2005].

3.1 Les deux principaux concepts

3.1.1 Le nom propre conceptuel

Pour une langue donnée, des noms propres totalement différents sur le plan graphique peuvent renvoyer à un même et unique référent et que ce phénomène se retrouve généralement d'une langue à l'autre. C'est le cas, par exemple, des noms propres *Jean-Paul II* et *Karol Józef Wojtyła* en français, des noms propres *John Paul II* et *Karol Józef Wojtyła* en anglais, etc. qui désignent tous la même personne mais à travers différentes langues.

Or, si le nom propre *Jean-Paul II* apparaît dans les lignes d'un quotidien français, un système de traduction automatique ne devra pas traduire ce nom propre en espagnol par *Karol Józef Wojtyła*, mais devra plutôt le traduire par le nom propre *Juan Pablo II*. Les noms propres *Jean-Paul II* et *Karol Józef Wojtyła* en français correspondent tous les deux à un certain point de vue sur un même et unique référent.

Nous définissons le nom propre conceptuel non pas comme le référent mais plutôt comme un certain point de vue sur celui-ci. Ainsi les noms propres *Allemagne* en français, *Alemania* en espagnol, *Deutschland* en allemand, etc. seront associés à un même nom propre conceptuel, tandis que les noms propres *République fédérale d'Allemagne* en français, *República Federal de Alemania* en espagnol, *Bundesrepublik Deutschland* en allemand, etc. seront associés à un autre nom propre conceptuel. Ces deux noms propres conceptuels seront en relation de synonymie.

Pour définir ces différents points de vue, nous nous sommes basés sur un marquage diasystématique, qui provient des travaux sur la métalexigraphie de [Coseriu, 1998] et de [Blanco, 2001].

[Coseriu, 1998] propose un diasystème basé essentiellement sur quatre variétés distinctes (figure 3.1) :

On constate, dans chaque état de langue (c'est-à-dire, même en faisant abstraction du développement de cette langue dans le temps) trois types fondamentaux de variétés (à savoir : la variété dans l'espace, la variété relative à la stratification socio-culturelle de la communauté parlante et la variété concernant les occasions, circonstances et finalités de l'emploi de la langue dans le discours).

DIA-S	DEFINITION
Diachronique	variété dans le temps
Diatopique	variété dans l'espace
Diastratique	variété relative à la stratification socio-culturelle
Diaphasique	variété concernant les finalités de l'emploi

FIG. 3.1 – Le diasystème de Coseriu.

Pour [Blanco, 2001], la marque diasystématique, associée à chaque unité lexicale, constitue une information indispensable qu'un dictionnaire électronique doit contenir. Dans son dictionnaire électronique de l'espagnol de langue générale, il distingue onze étiquettes différentes (figure 3.2).

La variation est un phénomène linguistique qui se retrouve dans toutes les langues et permet à une même entité d'apparaître sous des noms différents.

DIASYSTÈME	MARQUES
diastratique	soutenu, familier, vulgaire
diatopique	américanisme, dialectal
diachronique	vieilli, néologisme
diaintégratif	latinisme, argot
dianormatif	incorrect
diakonnotatif	péjoratif, enfantin
diamédiatique	oral, écrit
diaphasique	formel, informel
diatextuel	journalistique, administratif, littéraire
diatechnique	langue spécialisée
diafréquence	rare

FIG. 3.2 – Le diasystème de Blanco.

Le nom propre conceptuel nous servira de pivot entre différentes langues. Les noms propres associés à un même nom propre conceptuel seront considérés comme des traductions possibles d’une langue à l’autre.

Le nom propre conceptuel sera représenté dans notre modèle par un numéro d’identité unique (ID), le pivot.

Nous avons décidé de ne pas prendre comme pivot le référent d’un nom propre, car cela risque de poser un certain nombre de problèmes qui ne sont pas évidents à résoudre. Prenons par exemple le nom propre *Paris* et le nom propre *Lutèce*. Supposons que ces deux noms propres correspondent à un même référent. Si nous ajoutons une relation entre le pivot correspondant au nom propre *Paris* et au nom propre *France*, indiquant que *Paris* est la capitale de la *France*. Cela impliquera que *Lutèce* qui est relié au même pivot que *Paris* soit aussi la capitale de la *France*, ce qui n’est pas vrai. Nous serions donc obligés de déplacer cette relation au niveau linguistique. Au lieu d’avoir une relation unique au niveau indépendant des langues, nous aurions autant de relations qu’il y aurait de langues. De plus, la notion de référent est une notion extra-linguistique difficile à définir. Il n’est pas toujours évident d’associer un unique nom propre à un référent¹. Est-ce que l’on peut dire que le nom propre *France* et le nom propre *Gaule* correspondent à un même référent ? On sait que la *Gaule* ne correspondait pas surfaciquement à la *France* car elle comprenait la France actuelle, le nord de l’Italie et la Belgique. Pouvons-nous aussi dire que la Grèce antique et la Grèce d’aujourd’hui ont le même référent ?

3.1.2 Le prolexème

Dans notre modèle, le prolexème correspond à une projection du nom propre conceptuel dans une langue donnée. Chaque prolexème d’une langue donnée sera donc relié à un seul et unique nom propre conceptuel. C’est en se basant sur cette relation que l’on va pouvoir traduire les prolexèmes d’une langue vers une autre. Le concept de prolexème peut aussi se définir comme une classe d’équivalence de synonymes. Pour simplifier, nous considérons aussi le prolexème comme le lemme associé aux différentes formes d’un nom propre qui apparaissent dans les différents textes d’une langue donnée. Il peut ainsi être considéré comme la forme vedette d’un ensemble de dérivés et d’alias.

¹Une solution serait peut-être de considérer que le référent est le synset de nos noms propres conceptuels. Mais contrairement à WordNet, nous ne modélisons pas cette notion de synset.

Par exemple, les noms propres *Nations Unies*, *Onusien*, *ONU* auront *Organisation des Nations Unies* comme prolexème pour la langue française. Les noms propres *Onusian*, *United Nations* et *UNO* auront pour prolexème *United Nations Organization* pour la langue anglaise. Le prolexème français *Organisation des Nations Unies* et le prolexème anglais *United Nations Organization* seront reliés à un même nom propre conceptuel.

Les noms propres polysèmes, qui sont classés sous des catégories différentes, seront reliés à des prolexèmes différents. Par exemple, *Verdun* est à la fois connu comme étant une célèbre bataille durant la Première Guerre Mondiale, comme un traité entre les trois fils de l'empereur Louis le Pieux pour partager son Empire et, enfin, comme le chef-lieu de la Meuse. Pour ce cas-là, nous serons amené à créer trois prolexèmes différents associés à trois types différents. Par contre, dans le cas de toponymes correspondant à la fois à un lieu et une entité administrative (comme par exemple *Paris* qui est à la fois une ville et un département), nous avons décidé de ne pas dupliquer les prolexèmes pour éviter l'abondance d'homographes [Piton and Maurel, 2004]. Cette information sera rajoutée au niveau des expansions classifiantes du prolexème (voir section 3.3.4 page 69).

Les noms propres homographes seront aussi associés à des prolexèmes différents. En recherchant le nom propre *Sydney* dans un dictionnaire, on trouvera deux entrées distinctes : une qui correspondra à une ville en Australie et l'autre à une ville située au Canada. Il est à noter que l'homonymie dépend de la langue. Par exemple, en anglais, le nom propre *London* correspond à une ville du Canada ou à une ville en Angleterre, ce qui n'est pas le cas en français à cause de l'existence d'un exonyme (*Londres*), c'est-à-dire selon [ONU, 1977] :

Nom propre employé dans une certaine langue pour désigner un objet géographique situé à l'extérieur du territoire dans lequel cette langue a un statut officiel, et différent dans sa forme du nom propre utilisé dans la ou les langues officielles du territoire où l'objet géographique est situé.

3.2 Au niveau du prolexème

Les différents alias provenant d'un même nom propre peuvent être considérés comme des synonymes. Ainsi, les phrases (1), (2) et (3) sont sémantiquement identiques, puisque le nom propre *États-Unis d'Amérique* et ses deux alias, *États-Unis* et *USA*, renvoient à une même et unique entité.

- (1) *De leur côté, en effet, les **États-Unis** considèrent que les frontières doivent être le produit d'un accord. (Libération, le mercredi 24 mai 2006)*
- (2) *De leur côté, en effet, les **États-Unis d'Amérique** considèrent que les frontières doivent être le produit d'un accord.*
- (3) *De leur côté, en effet, les **USA** considèrent que les frontières doivent être le produit d'un accord.*

Certains dérivés d'un même nom propre peuvent aussi être considérés comme des synonymes de constructions contenant celui-ci, à la condition que le nom propre puisse être reconstruit en utilisant une transformation [Harris, 1968] [Harris, 1976]. Ainsi, le remplacement du dérivé *brésilien* dans la phrase (4) par le groupe prépositionnel contenant le nom propre *Brésil* (5) ne change pas le contenu sémantique de la phrase :

- (4) *Le président **brésilien** Luiz Inacio Lula da Silva a de fortes chances d'être réélu en octobre pour un second mandat. (Libération, le mardi 23 mai 2006)*
- (5) *Le président **du Brésil** Luiz Inacio Lula da Silva a de fortes chances d'être réélu en octobre pour un second mandat.*

Il existe bien sûr des cas où le dérivé n'est plus le synonyme exact du nom propre dont il dérive, car il n'existe pas de transformation qui les relie :

(6) *Petite firme familiale de distribution de lait **pasteurisé** établie dans les environs de Parme dans les années 1960...* (*Le Monde Diplomatique*, février 2004)²

(7) * *Petite firme familiale de distribution de lait de **Louis Pasteur** établie dans les environs de Parme dans les années 1960...*

(8) ?* *Petite firme familiale de distribution de lait **traité comme le préconise Louis Pasteur** établie dans les environs de Parme dans les années 1960...*

Dans cet exemple, la phrase (6) n'est plus sémantiquement équivalente à la phrase (7) et ni semble-t-il à la phrase (8), puisque dans la phrase (6) le dérivé lexicalisé *pasteurisé* signifie d'après le Petit Larousse :

Lait frais pasteurisé : lait frais ayant subi l'opération de pasteurisation par chauffage à une température de 72 à 85 °C pendant 15 à 20 secondes

Dans ce cas là, nous n'intégrerons pas ce dérivé dans notre dictionnaire électronique, puisque celui-ci relève plutôt d'un dictionnaire de noms communs.

Le classement des alias et des dérivés dans la partie qui dépend de la langue s'explique notamment par la raison que la créativité lexicale est propre à chaque langue. Une variante d'écriture existant dans une langue L_1 peut être totalement absente dans une langue L_2 . Un système de traduction automatique devra alors être capable de proposer une traduction de l'alias de la langue L_1 en utilisant la traduction du nom propre associé à cet alias dans L_1 . De même, pour le cas de la traduction d'un dérivé qui n'existe pas dans une langue donnée. Par exemple, le dérivé *Tourangeau* se traduira en anglais par *inhabitant of Tours*.

3.2.1 Les alias

Nous définissons les alias comme des synonymes qui dépendent de la langue. Nous avons regroupé dans le terme d'alias d'une part des synonymes exacts, les variantes d'écriture (caractères, abréviations, acronymes et sigles, transcriptions), les variantes orthographiques et d'autre part des synonymes approximatifs, diatopiques et diastratiques.

Il est parfois possible de définir des règles basées sur la structure interne [MacDonald, 1996] d'un prolexème afin de générer ses différents alias.

Variantes de caractères

La formation d'un alias résulte quelquefois de la variation d'un ou plusieurs des caractères qui composent le prolexème :

- la hauteur de casse : *Peugeot* ou *PEUGEOT*
- l'esperluette : *Science et Vie Junior* ou *Science & Vie Junior*³
- le remplacement des lettres comportant un signe diacritique : *Épinay-sur-Seine* ou *Epinay-sur-Seine* pour le français, *München* ou *Muenchen* pour l'allemand, *Århus*, *Arhus* ou *Aarhus* pour le danois
- le plus, le trait d'union et l'espace : *Canal Plus* ou *Canal +*
- l'ajout, le remplacement ou la suppression d'une ou plusieurs lettres : *Jean-François Delharpe* ou *Jean-François Delaharpe*
- etc.

²extrait du site <http://www.monde-diplomatique.fr/2004/02/RAMONET/10686> le 24/05/06

³Ces deux écritures ont été trouvées sur le site de <http://www.scienceetviejunior.fr> consulté le 23/05/06.

Variantes orthographiques

En français, la ligature n'est pas optionnelle le mot *cœur* ne doit pas s'écrire *coeur*, pour le cas des noms propres la lexicalisation n'est pas courante, dans les textes nous pouvons trouver les deux formes, nous avons décidé de les considérer comme des variantes. Il s'agit d'une variante orthographique provenant d'une erreur sur le diacritique (gluon) porté par le e de "cœur". Dans le cas des noms propres cette utilisation est moins stricte. Les deux noms *Crève-cœur-en-Brie* et *Crèvecoeur-en-Brie* se trouvent sur Google.

Abréviations

Les alias peuvent être aussi des abréviations du prolexème, c'est-à-dire :

Désignation formée par suppression de mots ou de lettres dans une forme plus longue désignant le même concept [ISO 1087-1 :2000].

Pour la plupart des noms de célébrités, il sera possible de modéliser la création de leurs alias en utilisant des règles basées sur leur structure interne. Par exemple, le nom propre *François Mitterrand* pourra être associé à la structure interne *Prénom Nom*, qui suit les règles d'aliasation suivantes :

- *Prénom Nom* -> *Prénom_ abrégé Nom* : *F. Mitterrand* (a)
- *Prénom Nom* -> *Nom* : *Mitterrand* (b)

Il existe malheureusement des exceptions parmi les noms de célébrités qui ne suivent pas ces règles. S'il est possible à partir de *François-René de Chateaubriand* de générer les alias *de Chateaubriand* et *Chateaubriand*, il ne sera par contre pas possible de créer l'alias *Gaulle* à partir de *Charles de Gaulle*. Il sera aussi moins évident d'appliquer la règle (a) à des noms propres non contemporains, comme *Marco Polo* (*?M. Polo) ou *Jules César* (*?J. César).

Nous pouvons aussi créer des règles pour les noms d'entreprise. A partir de *Sony Corporation*, on applique la règle *Nom Raison_ sociale* -> *Nom* pour créer l'alias *Sony* et la règle *Nom Raison_ sociale* -> *Nom Raison_ abrégée* pour obtenir l'alias *Sony Corp.*

De même, pour certains noms de villes françaises, il sera possible d'utiliser les deux règles suivantes :

- *Nom Préposition Hydronyme* -> *Nom* (c)
- *Nom Préposition Hydronyme* -> *Nom / Hydronyme* (d)

Ainsi, en appliquant la règle (c) et (d) sur la ville *La Roche sur Yon*, nous obtenons les deux alias suivant : *La Roche* et *La Roche / Yon*.

[Belleil, 1997], dans sa thèse, liste les principales abréviations (figure 3.3) qui peuvent s'appliquer sur les noms de lieu.

Acronymes et sigles

Les alias peuvent être des acronymes (*Sofres* pour *Société française d'enquêtes par sondages*) :

Abréviation formée des premières lettres des éléments constituant la forme complète de la désignation, ou des premières syllabes de la forme complète, et prononcée de façon syllabique [ISO 1087-1:2000].

Ils peuvent aussi être des sigles (*OCDE* pour *Organisation de coopération et de développement économiques*) :

Abréviation formée des premières lettres des éléments constituant la forme complète de la désignation et prononcée lettre par lettre [ISO 1087-1 :2000].

Lemme	Abréviation	Exemple
Saint	St	St-Cloud
Sainte	Ste	Ste-Hermine
Grand	Gd	Gd-Rullecourt
Grands	Gds	Les Gds-Chézeaux
Grande	Gde	La Gde-Verrière
Pont	Pt	Pt-Noyelles
Ponts	Pts	Saint-Denis-les-Pts
Mont	Mt	Dompierre-sur-Mt
Monts	Mts	Notre-Dame-de-Mts
sous	s	Clichy-s-Bois
sur	/	Vitry / Seine
-	espace	Dompierre sur Mt

FIG. 3.3 – Les variantes graphiques.

[Nakos, 1990] définit le sigle de la façon suivante :

nom (unité lexicale) formé d'initiales et de syllabes provenant : a) d'un mot (lexie simple), par exemple "T" pour "Titus" à l'époque romaine, b) d'un mot composé (lexie construite à partir d'au moins deux composantes), par exemple "ECG" pour "électrocardiogramme" ou c) d'un groupe de mots (lexie complexe), par exemple "CIA" pour "Central Intelligence Agency". Comme nous le savons déjà, le sigle se prononce soit lettre à lettre, par exemple "ECG" et "CIA", soit comme un mot, par exemple "UNESCO" ou "FNAC", plus difficile à prononcer ; dans ce dernier cas, il devient acronyme ou sigle acronymique.

La siglaison est un phénomène que l'on retrouve déjà à l'époque romaine. Par exemple, le sigle *INRI*, qui est inscrit sur certaines croix dans nos maisons et dans les églises signifie *Jesus Nazarenus Rex Iudaeorum*. C'est essentiellement à partir de la Seconde Guerre Mondiale, en raison de la création de nombreux organismes et d'entreprises multinationales, que le nombre de sigles s'est rapidement multiplié.

Certains sigles, tels que *UNESCO* (*Organisation des Nations unies pour l'éducation*) ou *OMS* (*Organisation mondiale de la santé*), sont des sigles internationaux. D'autres sont spécifiques à certaines langues, par exemple *OEA* (*Organisation des États américains*) pour le français et *OAS* (*Organization of American States*) pour l'anglais.

Certains sigles sont formés d'initiales provenant de noms composés d'une autre langue, comme par exemple en français *ESA* (*European Space Agency*) pour *Agence spatiale européenne*.

En français, les acronymes s'écrivent théoriquement toujours en majuscule ou avec une majuscule initiale (nouvelle écriture) et les sigles tout en majuscule avec des points ou sans point (nouvelle écriture). Certains acronymes et certains sigles sont aussi sujets à des variations sur certains des caractères qui les composent. Par exemple, on trouve deux écritures possibles pour l'acronyme *Association pour l'emploi dans l'industrie et le commerce* : *AS-SEDIC* ou *Assédic*. C'est le cas aussi pour l'*Organisation des nations unies* (*ONU*, *Onu* et *O.N.U.*) et l'*Institut National des Langues et Civilisations Orientales* (*INaLCO* et *IN-ALCO*).

Transcriptions

Nous avons aussi intégré les transcriptions et les translittérations dans les catégories d’alias. Une translittération est une opération qui consiste à transposer signe par signe un ou plusieurs mots écrits dans un système d’écriture vers un autre. Une transcription est souvent basée sur la phonétique.

Un même nom propre russe peut posséder en français plusieurs transcriptions différentes. Les transcriptions ne sont pas identiques d’une langue à l’autre. Voici un exemple⁴ avec le nom propre russe Владимир Владимирович Маяковский qui se transcrit :

- *Vladimir Vladimirovitch Mayakovski, Maïakovski*, ou *Mayakovsky* en français
- *Wladimir Wladimirowitsch Majakowski* en allemand
- *Vladimir Vladimirovich Mayakovsky* en anglais
- *Vladimir Vladimirovitsj Majakowski* en néerlandais
- *Vlagyimir Vlagyimirovics Majakovszkij* en hongrois

Le phénomène de transcription concerne aussi les noms propres chinois qui apparaissent souvent dans les textes journalistiques français. Actuellement, la transcription d’un nom chinois peut se faire suivant trois systèmes de transcription totalement différents : l’EFEO, le Wade-Giles et le pinyin. L’EFEO, système mis au point par l’École française d’Extrême-Orient, est basé sur l’alphabet latin et est très adapté pour les utilisateurs francophones. Le Wade-Giles est un système de transcription notamment utilisé dans la plupart des pays anglo-saxons. Enfin, le pinyin est devenu, depuis 1958, le système de transcription officiel adopté par le gouvernement chinois.

Les noms propres *Pékin*, *Tchang Kai-chek* et *Mao Tsé-toung* transcrits avec l’EFEO sont beaucoup plus connus des Français que leur forme pinyin *Beijing*, *Jiang Jieshi* et *Mao Zedong*.

En serbe, tous les mots, écrits en alphabet cyrillique, possèdent une transcription en alphabet latin. Par exemple, *Организација уједињених нација* en alphabet cyrillique se transcrit *Organizacija ujedinjenih nacija* (*Organisation des nations unies*) en alphabet latin.

Synonymes diastratiques

Certains alias peuvent être le résultat d’une transformation, d’un ajout ou d’une réduction : *Mère Angélique* (pour *Marie Jacqueline Angélique Arnaud*), *le Second Pitt* (pour *William Pitt*), etc.

Ces alias constituent des cas discutables. Nous avons décidé de les placer dans la partie qui dépend de la langue, car leur formation dépend souvent de la culture liée à une langue.

Synonymes diatopiques

Nous avons regroupé dans la variété diatopique les noms propres d’une ou plusieurs langues régionales d’un même pays qui sont en relation de synonymie avec le prolexème.

Certains pays présentent une grande diversité linguistique, notamment à travers la présence d’une ou plusieurs langues régionales. La France possède de nombreuses langues régionales telles que le catalan, le corse, le breton, le basque, etc. Par exemple, la ville de *Nantes* est appelée par les Bretons *Naoned*, qui est aussi un nom de la langue française. Il existe aussi quelques villes françaises qui possèdent un nom basque, par exemple, *Saint-Jean-de-Luz* qui donne en basque *Donibane Lohitzun*.

⁴Cet exemple a été trouvé sur le site http://fr.wikipedia.org/wiki/Vladimir_Mayakovski le 23/05/06

En français, *Naoned* sera considéré comme un synonyme diatopique de *Nantes*. En breton, *Naoned* serait le prolexème associé au prolexème français *Nantes*.

3.2.2 Les dérivés

Les dérivés de notre base sont considérés comme des synonymes, à une transformation près, des prolexèmes dont ils proviennent. Parmi les types de dérivés existant en français, nous avons principalement deux catégories : les noms relationnels et les adjectifs relationnels. [Daille, 1999] définit ainsi les adjectifs relationnels :

Les adjectifs relationnels possèdent les propriétés linguistiques suivantes : relation morphologique avec un nom, [...] possibilité d'équivalence avec un complément prépositionnel d'un nom de tête au sein d'un syntagme nominal (acidité sanguine, acidité du sang) [...] et un certain nombre d'autres propriétés comme le fait d'être inusités comme attribut, l'incompatibilité avec le degré, la postposition immédiate après le nom dans une séquence d'adjectifs postposés, etc.

Les noms relationnels en français débutent normalement par une majuscule (par exemple : *Parisien*, *Marseillais*, etc.). Leur classement au sein de la classe des noms propres est dû à leur identité référentielle fortement connotée. Cependant, cela ne se passe pas de la même manière dans toutes les langues.

Contrairement au nom relationnel, l'adjectif relationnel ne commence pas par une majuscule (par exemple : *parisien*, *marseillais*, etc.). A ce détail près, l'adjectif relationnel reste identique au nom relationnel, sauf quelques rares exceptions. Par exemple, *suisse* est l'adjectif relationnel féminin et *Suisse* est le nom relationnel féminin provenant du prolexème *Suisse*.

En français, la formation des dérivés à partir d'un prolexème ou d'un alias résulte de règles morphologiques complexes. Parfois, au lieu de s'appliquer à la base effective, ces règles s'appliquent à une forme supplétive. [Adouani, 1993] définit ce mécanisme ainsi :

Une paire de mots est ici dite supplétive si ses deux membres sont liés entre eux par une relation dérivationnelle dont la partie sémantique est régulière mais dont la partie formelle est, soit inexistante, soit profondément altérée. Cela peut être des mots complètement différents ou étymologiquement apparentés ou encore dont l'un a une forme latinisée ou réduite.

Les *Stéphanois* sont les habitants de la ville de *Saint-Étienne*. Étienne fut surnommé le couronné (en grec *stephanos*), car il fut le premier à recevoir la couronne de martyr.

Quand un prolexème appartient à la classe des toponymes, les noms relationnels qu'il engendre sont soit des gentilés (ou noms désignant les habitants d'une ville), soit des ethniques (noms désignant les habitants d'un pays ou d'une région). Selon [Eggert et al., 1998] et [Eggert, 2002], la création des gentilés en français est un phénomène très irrégulier. La figure 3.4 présente la répartition en pourcentage des suffixes apparaissant dans la formation d'un gentilé à partir de l'étude d'un corpus comprenant environ 2 757 gentilés.

En français, quelques noms de pays produisent des préfixes dérivés (*franco*, *américano*, etc.) provenant parfois d'une forme supplétive (*hispano*, *lusso*, etc.). Il arrive quelquefois aussi qu'un prolexème possède un nom relationnel diastratique comme dérivé. C'est le cas du prolexème *Paris* avec son dérivé *Parigot*. Cette forme est souvent connotée péjorativement.

Dans certaines langues, comme le serbe, les noms relationnels et les adjectifs relationnels ne sont pas systématiquement identiques et leur création se fait par des mécanismes morphologiques totalement différents du français. Le serbe, qui est une langue beaucoup

Suffixes	Pourcentage	Suffixes	Pourcentage
ois	36,1 %	aire	0,3 %
ais	25,1 %	iste	0,29 %
ien	18,7 %	at	0,29 %
éen	3,9 %	ar	0,22 %
in	3,3 %	asque	0,22 %
ain	3,2 %	enc	0,22 %
en	1,3 %	ol	0,22 %
on	0,9 %	ant	0,18 %
ard	0,7 %	and	0,15 %
ot	0,7 %	ate	0,11 %
an	0,6 %	ite	0,07 %
aud	0,4 %	iote	0,07 %
ier	0,4 %	autres	1,9 %
aux	0,4 %		

FIG. 3.4 – Statistique des suffixes de gentilés.

plus riche et complexe sur le plan morphologique, distingue deux catégories de noms relationnels : les noms relationnels féminins et les noms relationnels masculins. A partir de ces deux types de noms relationnels, il devient alors possible de former des adjectifs possessifs et des adjectifs relationnels. La formation des adjectifs ne se fait pas uniquement à partir des noms relationnels dérivés, mais peut aussi se faire à partir du prolexème ou des différents alias.

[Aljovic, 2000], dans sa thèse, explique la création des adjectifs possessifs de la façon suivante :

Les possessifs sont des éléments adjectivaux au même titre que les adjectifs lexicaux [...]. Le suffixe possessif ne s'attache qu'aux noms propres, ou aux substantifs utilisés comme des noms propres (à référence unique) [...]. Le suffixe s'attache de préférence aux noms au trait [+humain] [...]. Le suffixe possessif a deux formes -ov et -in. Le premier s'attache aux noms masculins, et le deuxième aux noms féminins.

A partir du prolexème *Београд* (Beograd en serbe latin, Belgrade en français) nous pouvons obtenir les dérivés suivants :

- београдски (beogradski) : adjectif relationnel (les rues belgradoises).
- Београдов (Beogradov) : adjectif possessif (l'allure de Belgrade).
- Београђанин (Beogradanin) : nom relationnel masculin (Belgradois).
 - београђански (beogradanski) : adjectif relationnel (les habitudes belgradois).
 - Београђанинов (Beogradaninov) : adjectif possessif (la maison d'un Belgradois)
- Београђанка (Beogradanka) : nom relationnel féminin (Belgradoise)
 - београђански (beogradanski) : adjectif relationnel (les habitudes des Belgradoises).
 - Београђанкин (Beogradankin) : adjectif possessif (la maison d'une Belgradoise)

3.3 Les relations

Une fois que les différents concepts du domaine des noms propres ont été identifiés et définis, il s'agit maintenant de rechercher les relations qui lient les noms propres entre eux. Dans cette partie, nous présenterons donc différentes relations linguistiques dans lesquelles peuvent intervenir des noms propres.

Les relations linguistiques qui peuvent exister entre les unités lexicales d'une langue sont essentiellement divisées en deux catégories distinctes [Polguère, 2003] : les relations paradigmatiques et les relations syntagmatiques. Lorsqu'une unité lexicale peut être substituée à une autre unité lexicale dans un même contexte, on dit alors que ces deux unités lexicales sont reliées par une relation paradigmatique. On distingue les relations paradigmatiques de similarité, comme la synonymie et l'antonymie, et les relations paradigmatiques d'inclusion, comme la méronymie et l'hyponymie. Les relations syntagmatiques sont des relations qu'entretiennent les unités lexicales entre elles dans une même phrase selon un principe de combinaison. Parmi les relations syntagmatiques, on trouve par exemple la relation de collocation.

Dans le cas des relations qui ne dépendent pas de la langue, nous avons retenu trois relations paradigmatiques (méronymie, synonymie, hyperonymie) et une relation syntagmatique (accessibilité). La relation d'hyponymie sera étudiée en détail dans le prochain chapitre, car elle nécessite l'introduction de la typologie des noms propres.

Dans le cas des relations qui dépendent de la langue, nous avons retenu deux relations syntagmatiques (l'expansion classifiante et l'éponymie).

3.3.1 La relation de synonymie

Le célèbre projet WordNet et ensuite le projet européen EuroWordNet ont fait de la relation de synonymie le pilier central sur lequel repose l'ensemble de leur architecture, notamment en la modélisant à travers le concept de "synset". Pour [Miller et al., 1990], la synonymie se définit de la façon suivante :

According to one definition (usually attributed to Leibniz) two expressions are synonymous if the substitution of one for the other never changes the truth value of a sentence in which the substitution is made. By that definition, true synonyms are rare, if they exist at all. A weakened version of this definition would make synonymy relative to a context : two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value.

Pour définir notre relation de synonymie, nous nous sommes inspirés de la définition que propose [Polguère, 2003], présentant la synonymie de la façon suivante :

La synonymie, c'est-à-dire l'identité de sens, est la relation lexicale sémantique par excellence :

Deux lexies L_1 et L_2 appartenant à la même partie du discours sont des synonymes exacts (ou synonymes absolus) si $(L_1) \equiv (L_2)$.

Ce sont des synonymes approximatifs si $(L_1) \cong (L_2)$. Dans ce dernier cas, il y a soit intersection soit inclusion de sens telle que L_1 et L_2 peuvent être considérées comme ayant une valeur sémantique suffisamment proche pour que l'une puisse être utilisée à la place de l'autre pour exprimer sensiblement la même chose.

Il est essentiel de noter que la synonymie lexicale exacte est rarissime.

Deux noms propres NP_1 et NP_2 appartenant à une même langue L peuvent être considérés comme étant en relation de synonymie si les trois conditions suivantes sont remplies :

- NP_1 et NP_2 correspondent tous les deux à un point de vue différent sur un même et unique référent, c'est-à-dire qu'il est possible de remplacer NP_1 par NP_2 (ou NP_2 par NP_1) pour exprimer sensiblement la même chose dans un contexte particulier.
- NP_1 ne peut se déduire de NP_2 , ni NP_2 de NP_1 , en appliquant des règles de morphologie dérivationnelle ou des règles d'aliasation (création d'alias), c'est-à-dire que NP_1 n'est pas un alias de NP_2 et NP_1 n'est pas un dérivé de NP_2 , et vice-versa.

Par exemple en français, le nom propre *Algérie* est en relation de synonymie avec le nom propre *République algérienne démocratique et populaire*. Ces deux noms propres se traduisent respectivement en roumain par *Algeria* et *Republica Democrată Algeria*.

Dans une synonymie, l'un des termes est souvent préférable à l'autre. On appellera le premier la forme canonique et l'autre la forme synonyme. Cette forme canonique en général correspond à la forme la plus connue. Par exemple, le nom propre *Molière* est plus connu que son synonyme *Jean-Baptiste Poquelin*.

Nous avons considéré la variation diatopique comme un alias (voir section 3.2.1 page 57). Il nous reste donc à présenter les trois variations restantes : diachronique, diastratique et diaphasique.

Diachronique

La première variation correspond à un point de vue diachronique, qui permet d'exprimer la notion de variété dans l'espace temporel, que l'on appelle aussi variation historique. Il s'agit principalement d'entités ou d'objets existants et connus sous un certain nom pendant une période donnée et qui, à cause de diverses raisons (politique, économique, stratégique, etc.), adoptent un nouveau nom et, à partir de cet instant, leur ancien nom cesse d'être utilisé. Voici un exemple⁵ de synonymie diachronique dans différentes langues :

- *Zaire* et *République démocratique du Congo* en français
- *Zaire* et *Democratic Republic of the Congo* en anglais
- *Zaire* et *República Democrática do Congo* en portugais
- *Zair* et *Demokratyczna Republika Konga* en polonais
- *Zaire* et *República Democrática del Congo* en espagnol
- *Zaire* et *Demokratische Republik Kongo* en allemand

Certaines transformations permettent aussi un passage entre ces synonymes, par exemple⁶ :

SUR LES MASSACRES DANS L'EX-ZAÏRE

[...]

DEPUIS l'arrivée au pouvoir de Mobutu Sese Seko en 1965, l'ancienne République démocratique du Congo (RDC), devenue Zaire...

(le Monde diplomatique)

Diastratique

La seconde variation est diastratique, c'est-à-dire liée à la classe socio-culturelle. Pour des raisons diverses (pseudonyme d'auteur, nom religieux, sobriquet, etc.), certains référents correspondent à plusieurs noms propres conceptuels. Nous avons rassemblé ici les variantes familières (*Paris* et *Paname*) et les variations savantes, dont voici un exemple dans plusieurs langues :

⁵La plupart de nos exemples ont été extraits du site <http://fr.wikipedia.org/>

⁶<http://www.monde-diplomatique.fr/1997/12/GARRETON/9657> consulté le 8/06/06

- *Jean-Paul II* et *Karol Józef Wojtyła* en français
- *John Paul II* et *Karol Józef Wojtyła* en anglais
- *Johannes Paul II.* et *Karol Józef Wojtyła* en allemand
- *Juan Pablo II* et *Karol Józef Wojtyła* en espagnol
- *Gioan Phaolô II* et *Karol Józef Wojtyła* en vietnamien

De même, lorsque l'on parle de l'auteur du roman *La mare au diable*, il sera préférable de parler de *George Sand* plutôt que de *Aurore Dupin, baronne de Dudevant* au risque d'être incompris par une majorité de personnes.

Diaphasique

Une variation diaphasique est liée à une différence de finalité d'emploi. Ainsi, pour un effet stylistique, on utilisera :

- *Paris* et *Ville lumière* en français
- *Paris* et *City of Light* en anglais
- *Paris* et *Stadt des Lichtes* en allemand
- *Parigi* et *Città delle luci* en italien
- *Paris* et *Cidade das Luzes* en portugais

Dans le cadre d'un article traitant de politique internationale, on considérera comme équivalent :

- *Allemagne* et *République fédérale d'Allemagne* en français
- *Alemania* et *República Federal de Alemania* en espagnol
- *Deutschland* et *Bundesrepublik Deutschland* en allemand
- *Germany* et *Federal Republic of Germany* en anglais
- *Niemcy* et *Republika Federalna Niemiec* en polonais

3.3.2 La relation de méronymie

La relation de méronymie constitue sans doute une des relations importantes que l'on retrouve dans le système WordNet et qui apparaît aussi dans de nombreux autres projets (EuroWordNet, Balkanet, SIMPLE, etc.). On l'appelle également relation partie-tout (*whole-part*), relation partitive ou encore relation d'inclusion. Lorsque deux unités lexicales A et B sont en relation de méronymie, on dit que A est un méronyme de B, et on dit que B est un holonyme de A si et seulement si A est une partie de B.

Cette relation permet d'établir une hiérarchisation sur plusieurs niveaux entre les éléments contenant (holonymes) et les éléments contenus (méronymes). Dans la plupart des cas, elle ne participe pas directement aux différentes étapes de la traduction, mais elle apporte une aide non négligeable dans le domaine de la recherche d'information.

[Winston et al., 1987], dans le cadre de leurs travaux, ont proposé un découpage de la relation de méronymie en six catégories différentes (figure 3.5).

Le projet européen EuroWordNet utilise ces six catégories de méronymie, tandis que WordNet se base uniquement sur trois catégories de méronymie :

- *part of* qui correspond à la catégorie *component/integral object*.
- *substance of* qui correspond à la catégorie *stuff/object*.
- *member of* qui correspond à la catégorie *member/collection*.

Dans le cadre de nos travaux sur les noms propres, seulement deux catégories de méronymes parmi les six catégories proposées par [Winston et al., 1987] s'appliquent :

- lieu/zone
- membre/collection

Type de méronymie	Exemple
component/integral object	handle/cup
member/collection	tree/forest
portion/mass	grain/salt
stuff/object	steel/bike
feature/activity	dating/adolescence
place/area	oasis/desert

FIG. 3.5 – Taxonomie des relations de méronymie.

Il faudrait aussi ajouter la relation de méronymie temporelle [Van Campenhoudt, 1996].

La figure 3.6 donne des exemples de relations de méronymie entre les différentes classes de noms propres.

3.3.3 La relation d’accessibilité

Les premiers travaux du projet Prolex, avant le début de cette thèse, portaient uniquement sur des toponymes. Il existait une relation qui permettait de préciser qu’une ville était la capitale d’une région ou d’un pays. Au début de nos études, nous avons étendu cette relation à la relation *Chef*, qui est une fonction lexicale appelée *Cap* du Dictionnaire Explicatif et Combinatoire du français contemporain. La relation *Chef* permettait de préciser si une entité est à la tête d’un groupe d’entité. En plus de la relation régions-capitales, nous pouvions modéliser les relations entre les anthroponymes et les anthroponymes collectifs. Mais très vite, nous avons voulu introduire une relation entre les auteurs et leurs œuvres, les personnes et leur famille, etc. C’est pour cela que nous avons décidé d’utiliser plutôt la relation d’accessibilité présentée par [Jonasson, 1994].

Si nous cherchons, par exemple, le nom commun *fourchette* dans un dictionnaire de langue française, on obtient le résultat suivant :

fourchette

nom féminin

1. *Ustensile de table à dents pointues, dont on se sert pour piquer les aliments.*

(Le Petit Larousse)

En cherchant le nom propre *Tours* dans le même dictionnaire, nous pouvons lire :

Tours [*tur*]

chef-lieu du département d’Indre-et-Loire, sur la Loire, à 225 km au S.-O. de Paris

(Le Petit Larousse)

En lisant la définition du nom commun *fourchette*, tout lecteur humain sera capable d’imaginer à quoi ressemble une fourchette et de la reconnaître dès qu’il la verra.

Dans le cas du nom propre *Tours*, il est difficile à partir de la définition que nous donne le dictionnaire de nous représenter la ville de Tours. Par contre, nous pouvons situer l’emplacement de cette ville sur une carte géographique. Dans le cas de la définition d’un personnage célèbre, nous aurions des informations sur le métier qu’il exerçait ou exerce, sur certains événements de sa vie, sa date de naissance, ses œuvres s’il s’agit d’un artiste, etc. Nous aurions dans ce cas tout l’historique de ce personnage célèbre, tandis que dans le cas d’un nom commun nous n’aurons pas d’information sur sa date d’apparition, son histoire, etc. Il s’agit d’une description encyclopédique de l’entité plutôt que d’une définition

<i>Type de méronymie</i>	<i>Exemple</i>
Célébrité / Association	François Hollande / le Parti Socialiste
Célébrité / Ensemble	Syd Barrett / les Pink Floyd
Célébrité / Entreprise	Franck Riboud / Danone
Célébrité / Institution	Marguerite Yourcenar / Académie française
Célébrité / Organisation	Shafqat Kakakhel / PNUE
Célébrité / Dynastie	Charlemagne / Carolingien
Célébrité / Œuvre	Lancelot du lac / Cycle du roi Arthur
Célébrité / Pays	Victor Hugo / France
Célébrité / Histoire	Louis XIV / Ancien Régime
Entreprise / Entreprise	Air France / Air France-KLM
Entreprise / Pays	SNCF / France
Entreprise / Supranational	EADS / Europe
...	...
Astronyme / Astronyme	Jupiter / le système solaire
Géonyme / Pays	la Forêt-Noire / Allemagne
Hydronyme / Pays	Seine / France
Hydronyme / Supranational	Danube / Europe
Pays / Supranational	France / Europe
Pays / Organisation	France / ONU
Région / Pays	La Vendée / France
Région / Région	Indre-et-Loire / Région Centre
Ville / Région	Tours / Indre-et-Loire
Ville / Œuvre	Minas Tirith / Le retour du roi
Édifice / Ville	Panthéon / Rome
Édifice / Œuvre	tour de Babel / Bible
Voie / Ville	la place de l'Étoile / Paris
...	...
Objet / Œuvre	Excalibur / cycle du roi Arthur
Œuvre / Œuvre	Le retour du roi / Le Seigneur des Anneaux
Produit / Produit	Mégane / Renault
...	...
Histoire / Histoire	la Prise de la Bastille / la Révolution française
...	...

FIG. 3.6 – Relation de méronymie entre les noms propres.

classique. Cette description relie le nom propre *Tours* à d'autres noms propres : *Indre-et-Loire*, *Loire*, *Paris*. En recherchant dans le dictionnaire le nom propre *Indre-et-Loire*, nous retrouvons une référence au nom propre *Tours*. Cependant, ce renvoi n'est pas systématique pour tous les noms propres du dictionnaire. Prenons par exemple le cas du nom propre *Aaron* dans le dictionnaire :

Aaron

XIIIe s. av. J.-C.

Frère aîné de Moïse et premier grand prêtre des Hébreux.

(Le Petit Larousse)

En lisant cet article, nous constatons une référence vers le nom propre *Moïse*. En recherchant l'article sur le nom propre *Moïse*, nous obtenons :

Moïse, en hébreu Moshé

XIIIe s. av. J.-C.

Libérateur et législateur d'Israël. La Bible le présente comme le chef charismatique qui a donné aux Hébreux leur patrie, leur religion et leur loi. Né en Égypte, il fut l'âme de la résistance à l'oppression que subissaient les Hébreux : il les fit sortir d'Égypte (l'Exode, vers 1250 av. J.-C.) et unit leurs divers groupes en un même peuple autour du culte de Yahvé. Il posa les éléments de base de la Loi (Torah).

(Le Petit Larousse)

Nous remarquons en lisant entièrement l'article que le nom propre *Aaron* n'y figure à aucun moment. En ironisant un peu, on pourrait affirmer que *Moïse* n'est pas le frère de *Aaron* et que la relation de fraternité entre ces deux personnes n'est pas une relation bijective. Cela montre bien que l'accès au nom propre *Aaron* se fait par l'intermédiaire du nom propre *Moïse* et non l'inverse.

A partir de ce constat, nous pouvons affirmer qu'un nom propre dans un dictionnaire n'est pas associé à une définition classique, mais à une description encyclopédique faite avec d'autres noms propres sur lesquels se base son accessibilité. Nous associons à chaque relation d'accessibilité un repérage (terme emprunté à [Jonasson, 1994]). Celui-ci précise la relation. Nous avons listé les différents repérages qui peuvent apparaître dans le cadre d'une relation d'accessibilité :

- parent : les personnes et les membres de leur famille. *Marie* est la mère de *Jésus*, *Louis XIII le Juste* est le fils de *d'Henri IV*, etc.
- créateur : les auteurs et les œuvres. *Richard Wagner* est le compositeur de *l'Anneau du Nibelung*, *Victor Hugo* est l'auteur de *Ruy Blas*, etc.
- capitale : les toponymes et leurs capitales. *La Rochelle* est le chef-lieu de la *Charente-Maritime*, *Bangkok* est la capitale de la *Thaïlande*, etc.
- dirigeant politique : les hommes politiques et les pays. *Jacques Chirac* est le président de la *République française*, etc.
- dirigeant non politique : les dirigeants et les entreprises. *Franck Riboud* est le PDG du groupe *Danone*.
- fondateur : les fondateurs d'une association, d'un groupe, d'une entreprise, d'un parti, d'une institution, etc. *J. Escrivá de Balaguer* est le fondateur de *l'Opus Dei*, *Richelieu* est le fondateur de *l'Académie française*, etc.
- élève : les disciples et leurs maîtres. *Aristote* est le disciple de *Platon*, etc.
- siège : les entreprises, associations ou organisations et le toponyme correspondant au siège social. *Peugeot* est une firme *sochaliennne*, etc.

- locataire : les bâtiments officiels et les dirigeants. *Dominique de Villepin* est le locataire de *Matignon*, etc.
- etc.

3.3.4 La relation d'expansion classifiante

Cette relation, que l'on appelle aussi relation de classifieur [Jonasson, 1994] associe à chaque prolexème une expansion. Un nom propre apparaît régulièrement dans les textes journalistiques, quelle que soit la langue, accompagné d'expansions se trouvant soit à sa gauche, soit à sa droite. Toutes les expansions qui existent dans une langue ne se retrouvent pas forcément dans une autre langue. Par exemple, le français distingue l'expansion *rivière* et *fleuve* pour le nom d'un cours d'eau alors que l'anglais utilise seulement l'expansion *river*. La traduction des expansions peut parfois poser quelques problèmes. Par exemple, la traduction de *Rechtsanwalt Paul Bischof* (allemand) ne donne pas en français *Avocat Paul Bischof*, mais plutôt *Maître Paul Bischof*. Si l'expansion d'un nom propre est omise dans un texte, il est parfois nécessaire de la rétablir lors de la traduction de celui-ci, afin d'apporter un complément d'information au lecteur. Ainsi, le nom propre *la Loire* deviendrait en anglais *the Loire River*.

Nous avons prévu d'associer aux expansions classifiantes des liens vers des descriptions syntaxiques (grammaires locales [Gross, 1989]) ou sémantiques (les classes d'objets [Le Pesant and Mathieu-Colas, 1998], EuroWordNet, Framenet [Fillmore et al., 2003]).

Prenons l'exemple de l'expansion *écrivain*, on pourra lui associer :

1. la Frame *Text_creation* (figure 3.7)
2. la grammaire locale de la figure 3.8
3. le concept *writer* dont le numéro ILI est *06438760n* (figure 3.9)
4. la classe d'objet correspondante.

En français, quelques noms propres, tels que les toponymes, se construisent dans une phrase avec des prépositions locatives [Constant, 2003] : *à, dans, en, sur*, etc. Nous avons aussi envisagé d'intégrer ces informations sous forme de grammaires locales.

3.3.5 L'éponymie

La relation d'éponymie se compose de trois relations : l'antonomase, le figement et la terminologie. Contrairement aux autres relations, l'objectif de la relation d'éponymie est d'empêcher une reconnaissance abusive des noms propres dans des textes. Par exemple, la *loi Pasqua* devra être reconnue comme terme mais pas *Pasqua* tout seul.

L'antonomase

L'antonomase est une figure de rhétorique par laquelle un nom propre est remplacé par un nom commun ou inversement. Nous avons pris en compte uniquement, dans le cadre de la relation d'antonomase, les antonomases à partir d'un nom propre.

Un nom propre employé en tant qu'antonomase perd la plupart du temps, dans le cas du français, sa majuscule initiale, surtout quand le lien qui l'unit au nom propre originel tend à s'effacer :

une mégère = une femme violente
un bic = un stylo-bille
un kleenex = un mouchoir en papier

Text_creation

Definition:

An **Author** creates a **Text**, either written, such as a letter, or spoken, such as a speech, that contains meaningful linguistic tokens, and may have a particular **Addressee** in mind. The **Text** may include information about its topic, although the latter is not an FE in this frame.

I **PENNED** a letter concerning racism to Congress.

The brothers **SAID** not two words to each other.

DOT any notes you need below the line in red pen only.

FEs:

Core:

Author [Author]
Semantic Type
Sentient

The **Author** produces a particular **Text**.

Text [text]

The entity which results from the act of writing or speaking.

Michael **WROTE** a frame description.

Cybil wanted to **SPEAK** those three words.

FIG. 3.7 – Extrait de la Frame *Text_creation*.

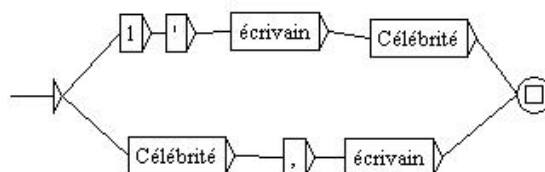


FIG. 3.8 – Exemple d'une grammaire locale.

Cette figure de style existe dans la plupart des langues :

pampersy pour *couches jetables* en polonais

kalodont pour *crème dentifrice* en serbe

biro pour *stylo-bille* en anglais

ксерокс (xerox) pour *photocopieuse* en russe

Certaines antonomases peuvent exister dans une langue et être totalement absentes dans d'autres langues. Le nom propre *Pampers* a donné lieu à une antonomase en polonais, alors que ce n'est pas le cas en français.

Le figement

Pour le dictionnaire anglais *Cobuild Dictionary of Idioms*, les idiomes sont définis de la façon suivante :

An idiom is a special kind of phrase. It is a group of words which have a different meaning when used together from the one it would have if the meaning of each word were taken individually. [...] Idioms are typically metaphorical : they are effectively metaphors which have become 'fixed' or 'fossilized'.

entity
 object, physical object
 living thing, animate thing
 organism, being
 person, individual, someone
 communicator
 writer

FIG. 3.9 – Hiérarchie de *writer* dans EuroWordNet.

Nous définissons le figement comme des tournures idiomatiques construites à partir d'un ou plusieurs noms propres. Certaines tournures idiomatiques comprenant un nom propre dans une langue donnée peuvent se traduire vers une autre langue à l'aide d'une autre tournure idiomatique pouvant ne pas comporter de nom propre. C'est le cas des exemples suivants :

être en tenue d'Adam = to be in one's birthday suit
not for all the tea in China = pour rien au monde
I don't know him from Adam = je ne le connais ni d'Ève ni d'Adam
 (Dictionnaire Hachette-Oxford)

Le sens d'un figement peut varier d'une langue vers une autre :

zwischen Scylla und Charybdis [sein] = être entre deux dangers
 (d'après Duden 11 / Redewendungen)
(tomber) de Charybde en Scylla = quitter un mal pour un autre pire encore
 (Petit Larousse)

La terminologie

On retrouve de nombreux noms propres dans les terminologies scientifiques (le *théorème d'al-Kashi* sur le calcul des longueurs des côtés d'un triangle non rectangle, les *équations de Maxwell* qui caractérisent les interactions entre charges, etc.), juridiques (la *loi Evin* relative à la lutte contre le tabagisme et l'alcoolisme, la *loi de Robien* sur l'investissement locatif, etc.) ou médicales (la *maladie de Creutzfeldt-Jakob*, la *maladie de Parkinson*, etc.).

Nous n'avons pas intégré ces termes dans la classe des noms propres, car ils appartiennent plus à une langue spécialisée qu'à la langue générale. De plus, leur traduction, loin d'être triviale, nécessite parfois l'utilisation d'une expression définie. Par exemple, la *loi Pasqua* ne se traduira pas en allemand par *Pasqua-Gesetz* mais plutôt par *französisches Einwanderungs- und Staatsangehörigkeitsgesetz*. Quelques fois, on sera amené à traduire un nom propre par son dérivé :

le théorème de Pythagore = der pythagoreische Lehrsatz
la maladie de Parkinson = die parkinsonsche Krankheit

D'une langue à l'autre, les noms propres utilisés dans une terminologie peuvent être sujets à des variations :

Comme nous l'avons montré, les noms propres utilisés dans les termes médicaux peuvent être composés, notamment reliés par un trait d'union, et l'ordre de composition des noms propres pour un même terme peut varier d'une langue à l'autre

("maladie de Legg-Perthes-Calvé" ; "Legg-Perthes-Calvé disease" ; "Perthes-Legg-Calvé-Krankheit").

[Bodenreider and Zweigenbaum, 2000b]

Finally, compound proper names found in different translations of ICD-10 sometimes show variation in the order or even in the number of the names (e.g. an alternate term for "relapsing panniculitis" is "Weber-Christian disease" in English, "maladie de Weber-Christian" in French, but "Pfeifer-Weber-Christian-Krankheit" in German).

[Bodenreider and Zweigenbaum, 2000a]

3.4 Représentation sous forme d'un schéma

Nous représentons les différents concepts du domaine des noms propres sous la forme d'une arborescence (figure 3.10) qui peut se décomposer en deux niveaux : un niveau indépendant des langues et un niveau dépendant de la langue.

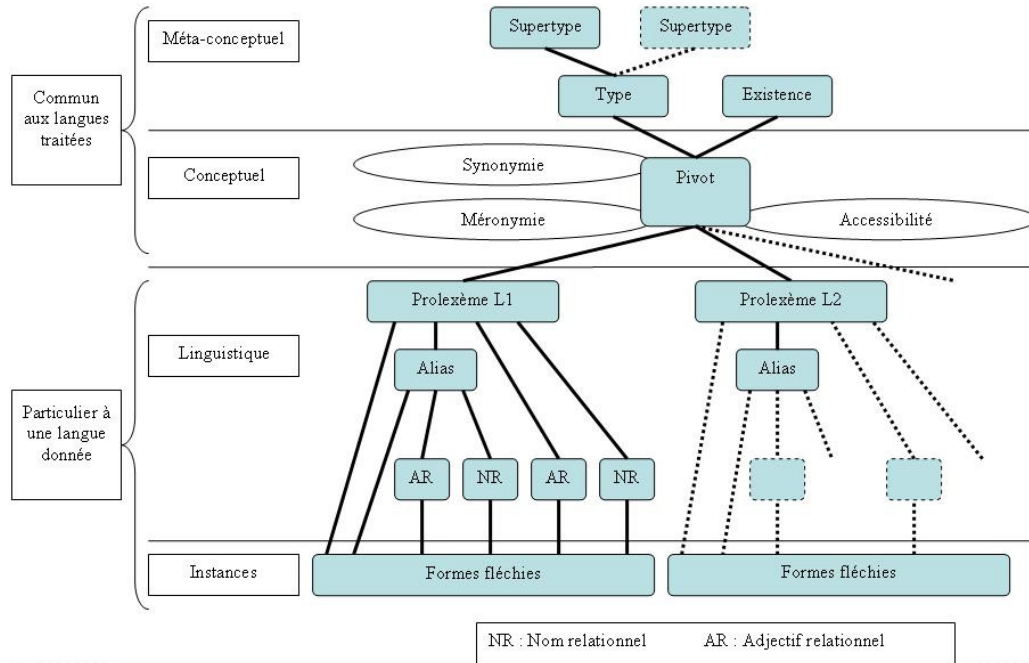


FIG. 3.10 – Les quatre niveaux.

Le niveau indépendant de la langue est lui-même composé de deux niveaux. Le premier niveau, que l'on appelle le niveau méta-conceptuel, comprend les types et l'existence (voir chapitre suivant). Le deuxième niveau est le niveau conceptuel, qui inclut le concept de nom propre conceptuel et les relations de méronymie, de synonymie et d'accessibilité.

Le niveau dépendant de la langue est aussi subdivisé en deux niveaux différents : le niveau linguistique et le niveau des instances. Le niveau linguistique englobe les concepts de prolexème, d'alias et de dérivé. Chaque langue possédera sa propre arborescence à partir de la forme canonique d'un nom propre, ou prolexème, qui sera relié à un même niveau indépendant de la langue à travers un ensemble de noms propres conceptuels. En raison des grandes divergences et de la complexité des mécanismes s'appliquant sur les noms propres,

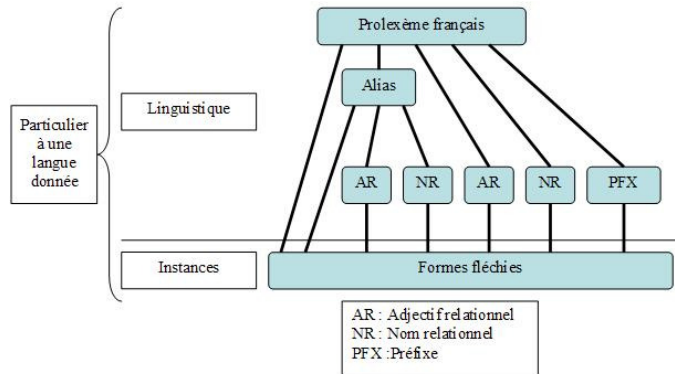


FIG. 3.11 – Le prolexème français.

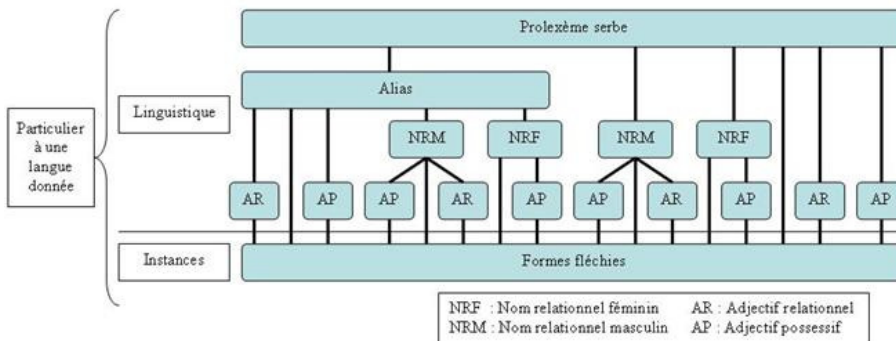


FIG. 3.12 – Le prolexème serbe.

nous ne pouvons définir une arborescence générale qui pourrait s'appliquer pour toutes les langues. Selon la langue, cette arborescence pourra être plus ou moins complexe. La figure 3.11 présente l'arborescence du prolexème français et la figure 3.12 l'arborescence plus complexe du prolexème serbe.

Le niveau des instances regroupe toutes les formes fléchies que l'on peut obtenir en appliquant des règles morphologiques, plus ou moins compliquées selon les langues, sur un nom propre. L'ensemble de ces formes fléchies, qui correspondent aux mots que l'on retrouve dans un texte, forme ce que [Polguère, 2003] appelle les lexies d'un nom propre :

Une lexie, aussi appelée unité lexicale, est un regroupement 1) de mots-formes ou 2) de constructions linguistiques qui ne se distinguent que par la flexion.

La figure 3.13 montre l'exemple détaillé du nom propre *Belgrade* en cyrillique serbe. La figure 3.14 détaille les prolexèmes français *Suisse* et *Confédération helvétique*.

Conclusion

Dans ce chapitre, nous avons défini nos deux principaux concepts : le nom propre conceptuel et le prolexème. Autour de ces deux concepts, nous avons ajouté d'autres concepts et défini des relations que nous avons représentés sous la forme d'un graphe divisé en deux niveaux (un niveau indépendant de la langue et un niveau dépendant de la langue).

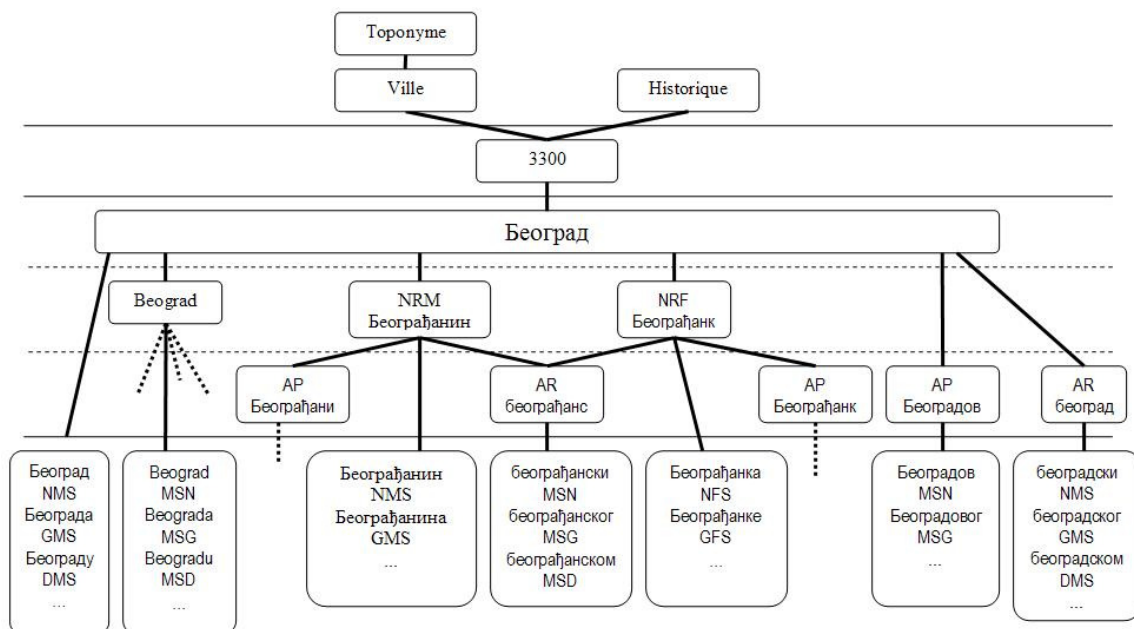


FIG. 3.13 – La partie cyrillique du prolexème serbe *Belgrade*.

Le modèle que nous présentons dans cette partie est donc le résultat de plusieurs modèles successifs. Nous espérons que la comparaison avec d'autres langues permettra de le valider définitivement.

Un essai avec le coréen a été concluant et nous a permis d'intégrer à Prolexbase une première langue non indo-européenne.

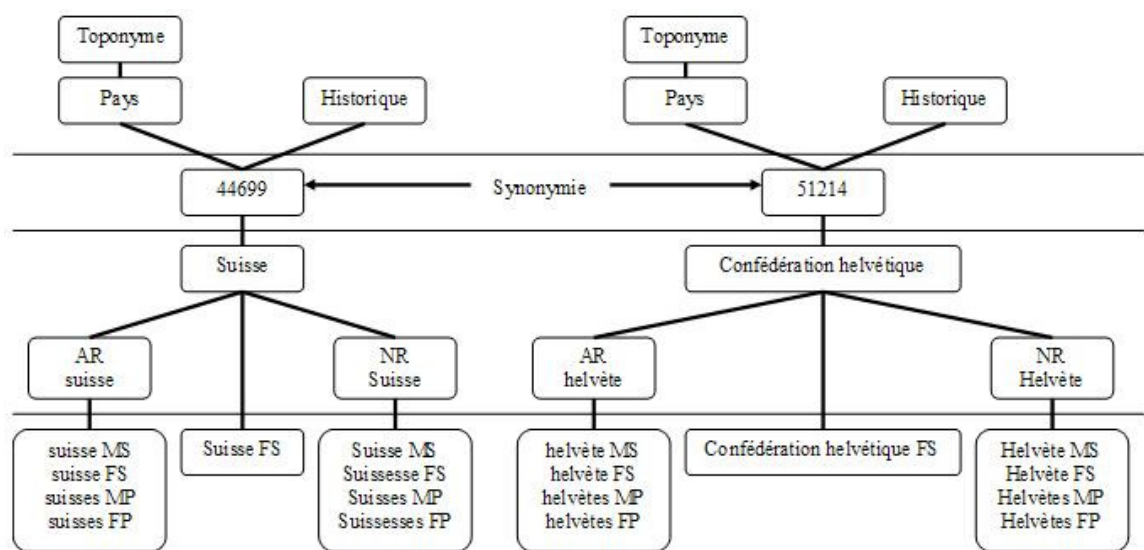


FIG. 3.14 – Les prolexèmes français *Suisse* et *Confédération helvétique*.